1   ANNETTE L. HURST (SBN 148738)
     ahurst@orrick.com
2   DANIEL D. JUSTICE (SBN 291907)
     djustice@orrick.com
3   ORRICK, HERRINGTON & SUTCLIFFE LLP
     405 Howard Street
4   San Francisco, CA 94105-2669
     Telephone: +1 415 773 5700
5   Facsimile: +1 415 773 5759

6   WILLIAM W. OXLEY (SBN 136793)
     woxley@orrick.com
7   ALYSSA CARIDIS (SBN 260103)
     acaridis@orrick.com
8   ORRICK, HERRINGTON & SUTCLIFFE LLP
     355 S. Grand Avenue
9   Los Angeles, CA 90071
     Telephone: +1 213 629 2020
10   Facsimile: +1 213 612 2499

11   *Attorneys for GitHub, Inc. and Microsoft Corporation*

12   UNITED STATES DISTRICT COURT

13   NORTHERN DISTRICT OF CALIFORNIA

14   OAKLAND DIVISION

15

16   J. DOE 1, et al.,   Case No. 4:22-cv-6823-JST

17   Individual and   Consolidated with Case No. 4:22-cv-7074-JST
            Representative Plaintiffs,
18   **DEFENDANTS GITHUB AND**
          v.   **MICROSOFT'S NOTICE OF MOTION**
19   **AND MOTION TO DISMISS PORTIONS**
     GITHUB, INC., et al.,   **OF THE SECOND AMENDED**
20   **COMPLAINT IN CONSOLIDATED**
          Defendants.   **ACTIONS**
21
                                            Date: May 16, 2024
22   ─────────────────────────────   Time: 2:00 p.m.
     AND CONSOLIDATED ACTION   Courtroom: 6, 2d Floor
23   Judge: Hon. Jon S. Tigar

24   Second Amended Complaint Filed: January
                                            24, 2024
25

26

27

28
                                                   GITHUB AND MICROSOFT'S MTD SECOND
                                                               AMENDED COMPL.
                                                              NO. 4:22-CV-6823-JST

TABLE OF CONTENTS

1

TABLE OF AUTHORITIES

2

Page(s)

3

Cases

iii

GitHub and Microsoft's MTD Second
Amended Compl.
No. 4:22-cv-6823-JST

NOTICE OF MOTION AND MOTION

TO PLAINTIFFS AND THEIR ATTORNEYS OF RECORD:

PLEASE TAKE NOTICE THAT on May 16, 2024 at 2:00 p.m., before the Honorable Jon S. Tigar seated in Courtroom 6 of the United States Courthouse at Oakland, California, with appearances to be made by Zoom videoconference unless otherwise ordered by the Court, Defendants GitHub, Inc. ("GitHub") and Microsoft Corporation ("Microsoft") will, and hereby do, move, pursuant to Federal Rule of Civil Procedure 12(b)(6) to dismiss with prejudice portions of the operative Second Amended Complaint in these consolidated actions, filed as ECF No. 198 in the 4:22-cv-6823 action (the "Second Amended Complaint" or "SAC"), as to both GitHub and Microsoft.

The grounds for the Motion are as follows. Plaintiffs again attempt to replead claims under § 1202(b)(1) and (b)(3) of the Digital Millennium Copyright Act ("DMCA"), but this Court should again dismiss those claims pursuant to Rule 12(b)(6). Plaintiffs fail to state a claim under § 1202 because the factual allegations required to support necessary elements are missing or implausible. Additionally, Plaintiffs' requests for unjust enrichment and punitive damages should be dismissed as improper and unsupported. Because Plaintiffs have already had three opportunities to plead these claims, they should be dismissed without further leave to amend.

1

MEMORANDUM OF POINTS AND AUTHORITIES

2

INTRODUCTION AND SUMMARY OF ISSUES

3   After multiple rounds of pleading, little is left of Plaintiffs' claims of "software piracy on

4   an unprecedented scale." SAC ¶ 202. This Court most recently granted GitHub and Microsoft's

5   motions to dismiss Plaintiffs' claims under the DMCA, because Plaintiffs at best alleged the

6   possibility that Copilot might output variations from code in their repository, and thus failed to

7   meet § 1202(b)'s requirement of an identical copy. ECF No. 189 at 15-16. Following that

8   decision, Plaintiffs have dropped any entitlement to statutory damages under the DMCA,

9   acknowledging they cannot allege that Copilot has ever done anything that violates § 1202. The

10 latest iteration of their complaint attempts to allege that Copilot *might* one day do what Plaintiffs,

11 despite considerable effort, have failed to show it would. This attempt fails.

12   In support of this theory, the SAC adds new assertions attempting to bolster an allegation

13 in the prior complaint—namely, that Copilot has a filter to block "verbatim excerpts of 'about

14 150 characters.'" SAC ¶ 145; *see* First Amended Complaint ("FAC") ¶ 133. The SAC now

15 alleges that a feature—called a "duplication detection" tool—allows users to check output against

16 public repositories for matches of about 150 characters and, if desired, to block such matches.

17 SAC ¶ 146. If a user has the automatic block feature turned off, then a coordinated "code

18 referencing" tool allows users to see information about code in repositories matching the snippet.

19 SAC ¶ 147. The SAC's newfound reliance on GitHub's duplication-detection and code-

20 referencing features is ironic, to say the least. Having spent the past 16 months claiming that they

21 are not getting the required credit for their work, Plaintiffs now complain about a set of tools that

22 provide just such attribution.

23   More importantly, this shift reflects an abandonment by Plaintiffs of their attempts to get

24 Copilot to spit out verbatim copies of their code. This abandonment crystallizes another

25 fundamental defect in the § 1202 claim: Plaintiffs do not, and seemingly no longer even attempt

26 to, allege removal of CMI from *a work*. Rather, their bid at showing plausibility based on the

27 duplication-detection and code-referencing features reflects the possibility of potential matches of

28 "about 150 characters" of someone's code, with no indication of any such outputs approaching

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

1   the length of any copyrighted work. To get a sense of what 150 characters looks like, the

2   sentence that precedes this one is longer. With the § 1202 claim shrunken to a "maybe-snippets"

3   theory, the SAC no longer alleges that Copilot removes CMI from any *work* at all—as

4   § 1202(b)(1) or (b)(3) require. And certainly the SAC does not *identify* any work in which

5   Plaintiffs own a copyright interest and from which CMI has been or may be removed. So, even

6   before getting to Plaintiffs' attempts to satisfy the identicality requirement, their claims fail for

7   lack of identification of any copyrightable work. *Infra* § I.A.

8   When it does come to identicality, the SAC fails to cure the problem that previously led

9   this Court to dismiss the DMCA claim. As with the First Amended Complaint, the current

10 version continues to acknowledge that "more often" Copilot's suggestions are "a modification."

11 SAC ¶ 108. It also continues to rely on the assertion that 1% of Copilot output may match *some*

12 code—a statement that most certainly does not, as Plaintiffs insinuate, mean that there is a 1%

13 chance that the output would ever match *their* code. Added to these insufficient contentions are

14 two sets of allegations that do not move the needle at all. One set involves the "duplication

15 detection" feature noted above. This makes little sense—a duplication-detection feature makes

16 use of duplicate outputs *less* likely. But certainly the mere existence of the feature does not make

17 an identical copy of *Plaintiffs*' code *more* likely, particularly when Plaintiffs' own elaborate

18 efforts to generate such a copy have failed, as the last round of briefing showed. As for the

19 second set of allegations, Plaintiffs rely on an academic paper that says nothing more than what

20 Plaintiffs have already alleged—that Copilot rarely, when elaborately prompted, might emit a

21 matching code snippet. The SAC still does not offer any reason to believe that *Plaintiffs'* code, as

22 opposed to all others, would be generated among those 1% of suggestions. Again, Plaintiffs have

23 tried, and now concede they failed, to get Copilot to do this. The academic papers they cite get

24 the SAC nowhere near alleging a plausible claim. *Infra* § I.B.

25   The SAC's maybe-snippet theory based on duplication-detection and code-referencing

26 tools has also forced the § 1202 claim even further from the mark: It is now entirely implausible,

27 from the face of the Complaint, that Copilot would ever make infringement of supposedly

28 protectable aspects of Plaintiffs' works easier to commit or harder to detect. Such a glaring

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-cv-6823-JST

1   deficiency merits dismissal even at the pleading stage under *Stevens v. Corelogic, Inc.*, 899 F.3d

2   666 (9th Cir. 2018). *Infra* § I.C.

3   Finally, Plaintiffs' requests for monetary relief in the form of unjust enrichment and

4   punitive damages are unsupported by any pleadings in the Second Amended Complaint. This

5   Court previously dismissed Plaintiffs' unjust enrichment claim with prejudice, ECF No. 189 at

6   11-13, and should now dismiss their prayer for unjust enrichment , as well as the request for

7   punitive damages. *Infra* § II.

8   **FACTUAL AND PROCEDURAL BACKGROUND**

9   GitHub and Microsoft briefly restate the relevant background only as relevant to the issues

10   in this Motion.

11   **A.   The Technology At Issue – According To The Second Amended Complaint.**

12   **1.   OpenAI Develops A Generative AI Tool Called Codex.**

13   OpenAI is a nonprofit organization that develops machine learning models, also referred

14   to as "Artificial Intelligence." SAC ¶¶ 2, 188, 190. Such models are typically trained through

15   exposure to a corpus of material called "training data." SAC ¶ 93. The patterns discerned from

16   the set of training data become part of the model, which can then generate answers based upon

17   those patterns in response to user prompts. SAC ¶ 93; *see* SAC ¶ 64.

18   The models at issue embody "inferred … statistical patterns governing the structure of

19   code," SAC ¶ 64, which they have discerned from the training data based on "a complex

20   probabilistic process." SAC ¶ 91. They are thus capable, in response to a prompt, of

21   "predic[ting] … the most likely [coding] solution." SAC ¶ 91. The models "return[] the solution

22   … found in the most [coding] projects when those projects are somehow weighted to adjust for

23   whatever variables [the model] ha[s] identified as relevant." SAC ¶ 91.

24   **2.   GitHub Offers Copilot, A Code Completion Tool Based On Generative AI.**

25   GitHub Copilot is a programming assistant. SAC ¶ 8. The Second Amended Complaint

26   alleges that Copilot "uses the OpenAI Codex to suggest code and entire functions in real-time" to

27   software developers. SAC ¶ 59.

28   Plaintiffs allege that the generative AI models that power Copilot were trained on billions

1   of lines of code that GitHub users stored in public GitHub repositories. *See* SAC ¶¶ 95, 210.

2   When GitHub users put their code on GitHub, they choose whether to make the code repositories

3   private or public. SAC ¶ 181. Users who set their repositories "to be viewed publicly … grant

4   each User of GitHub a nonexclusive, worldwide license to use, display, and perform Your

5   Content through the GitHub Service and to reproduce Your Content solely on GitHub as

6   permitted through GitHub's functionality." SAC Ex. 1 at 27 (GitHub Terms of Service ("TOS")

7   at 7). Every user agrees to GitHub's TOS, which include a "License Grant" to GitHub to "store,

8   archive, parse, and display … and make incidental copies" as well as "parse it into a search index

9   or otherwise analyze it" and "share" the content in public repositories with other users. SAC Ex.

10 1 at 26-27 (GitHub TOS at 6-7). And users can also select from a range of preset open source

11 licenses to apply to the code published in their various GitHub repositories, apply their own

12 individual licenses, or select none at all. SAC ¶ 46 n.4 & Appx. A.

13   Any GitHub user thus appreciates that code placed in a public repository is genuinely

14 public. Anyone is free to examine, learn from, and understand that code, as well as repurpose it

15 in various ways. And, consistent with this open source ethic, neither GitHub's TOS nor any of

16 the common open source licenses prohibit either humans or computers from reading and learning

17 from publicly available code. *See* SAC ¶ 46 n.4 & Appx. A.

18 **B.   Prior Proceedings.**

19   **1.   Plaintiffs' Original Complaint**

20   In the initial Complaint, Plaintiffs raised twelve claims against Defendants. Compl., *Doe*

21 *3 v. GitHub, Inc.*, No. 22-cv-7074-JST (N.D. Cal. Nov. 10, 2022), ECF No. 1. GitHub and

22 Microsoft moved to dismiss the entirety of the initial operative complaint pursuant to Federal

23 Rules of Civil Procedure 8, 9, 10, 12(b)(1), and 12(b)(6). Mots. To Dismiss Operative Compl. in

24 Consol. Actions, ECF No. 50.

25   This Court granted in part and denied in part GitHub and Microsoft's motions to dismiss.

26 ECF No. 95. On standing, this Court held that Plaintiffs "failed to establish an injury-in-fact

27 sufficient to confer standing for their claims for damages based on injury to property rights." *Id.*

28 at 10. It recognized, *id.* at 6, that "at the pleading stage, the plaintiff must 'clearly … allege facts

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

1   demonstrating each element'" of standing—(1) a personal injury in fact; (2) that the injury was

2   likely caused by the defendant, and (3) that the injury would likely be redressed by judicial relief.

3   *Spokeo, Inc. v. Robins*, 578 U.S. 330, 338 (2016) (quoting *Warth v. Seldin*, 422 U.S. 490, 518

4   (1975)). And it found that Plaintiffs had failed to demonstrate that "they themselves" had

5   suffered injury from "Copilot 5cenes5ce[ing] Plaintiffs' code as output with missing or incorrect

6   attribution, copyright notices, and license terms." ECF No. 95 at 8. Plaintiffs therefore lacked

7   standing to pursue retrospective relief (*i.e.*, damages). This Court also held that Plaintiffs failed

8   to allege facts demonstrating an injury-in-fact sufficient to confer standing for their privacy-based

9   claims. *Id.* at 7. Nevertheless, this Court held that "Plaintiffs plausibly allege that there is at least

10   a substantial risk that Defendants' programs will produce Plaintiffs' licensed code as output," and

11   that this risk "may support standing for injunctive relief." *Id.* at 9.

12   On the merits, this Court dismissed Plaintiffs' claims for violation of § 1202(a) and

13   § 1202(b)(2) of the DMCA, tortious interference, fraud, false designation of origin, violation of

14   the CCPA, unjust enrichment, civil conspiracy, declaratory relief, and various aspects of

15   Plaintiffs' derivative UCL claims for failure to state a claim under Rule 12(b)(6). *Id.* at 16. The

16   Court denied Defendants' motions to dismiss the breach of contract claim, ECF No. 95 at 22, as

17   well as Plaintiffs' claim based on § 1202(b) of the DMCA. The Court gave Plaintiffs leave to

18   amend as to most of the dismissed claims.

19   **2.   Plaintiffs' First Amended Complaint**

20   In their First Amended Complaint, Plaintiffs joined a fifth J. Doe GitHub user as a new

21   Plaintiff and alleged eight claims for relief. FAC ¶¶ 23; 183-289. The allegations underlying

22   Plaintiffs' claims under § 1202(b)(1) and (3), FAC ¶¶ 183-213 (Count 1), and for breach of open

23   source licenses, FAC ¶¶ 214-29 (Count 2), were largely unchanged from the initial Complaint.

24   Plaintiffs also included claims for breach of contract for selling licensed materials in violation of

25   GitHub's policies, FAC ¶¶ 230-40 (Count 3); intentional interference with prospective economic

26   relations, FAC ¶¶ 241-53 (Count 4); negligent interference with prospective economic relations,

27   FAC ¶¶ 254-65 (Count 5); a modified unjust enrichment claim, FAC ¶¶ 266-74 (Count 6); a

28   modified unfair competition claim, FAC ¶¶ 275-81 (Count 7); and a negligence claim, FAC

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

1   ¶¶ 282-89 (Count 8).

2   Plaintiffs continued to allege that they "published Licensed Materials they owned a

3   copyright interest in to at least one GitHub repository under one of the Suggested licenses." FAC

4   ¶¶ 19-23. Plaintiffs alleged that because Copilot was trained on public GitHub repositories, they

5   can be "reasonably certain" their Licensed Materials were "ingested by Copilot and [are]

6   sometimes returned to users as Output." FAC ¶ 97.

7   New to the First Amended Complaint, however, was Plaintiffs' acknowledgement that

8   "more often" than a snippet reproducing Plaintiffs' code are snippets with "variations" from

9   existing code. FAC ¶ 96. And in an apparent attempt to gain standing to seek damages, three of

10   the Plaintiffs also alleged, with respect to code in their repositories, that they were able to

11   "prompt[] Copilot" to "emit[]" their own Licensed Materials. FAC ¶¶ 98, 100, 103, 110, 123,

12   125, 127-28. These examples only illustrated, however, instances where Copilot suggestions

13   differed from code contained in Plaintiffs' repositories.

14   Take Plaintiffs' example regarding Doe 1. To generate it, Plaintiffs entered, character for

15   character, 22 lines of Doe 1's code (contained in the boxes in the examples below) to yield a

16   suggested output of 18 additional lines of code. FAC ¶¶ 106-08. Even then Plaintiffs conceded

17   the output "is not an exact match for Doe 1's code." FAC ¶ 108.

18

19

20

21

22

23

24

25

26

27

28

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

Original code:

Output:

7
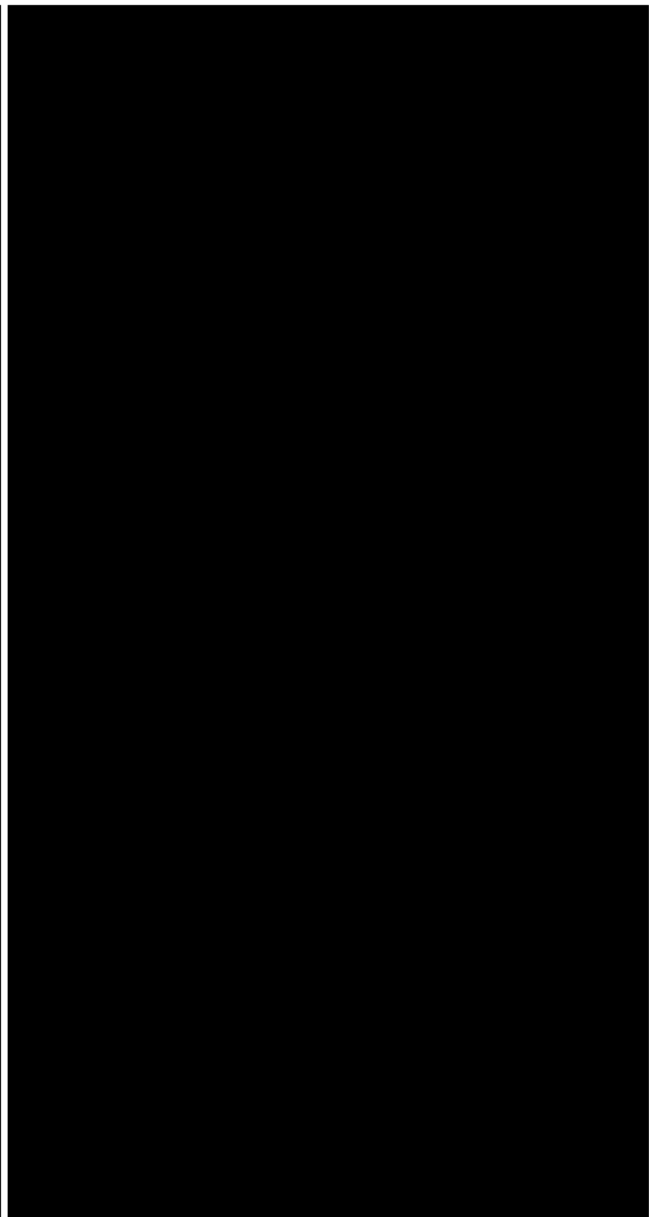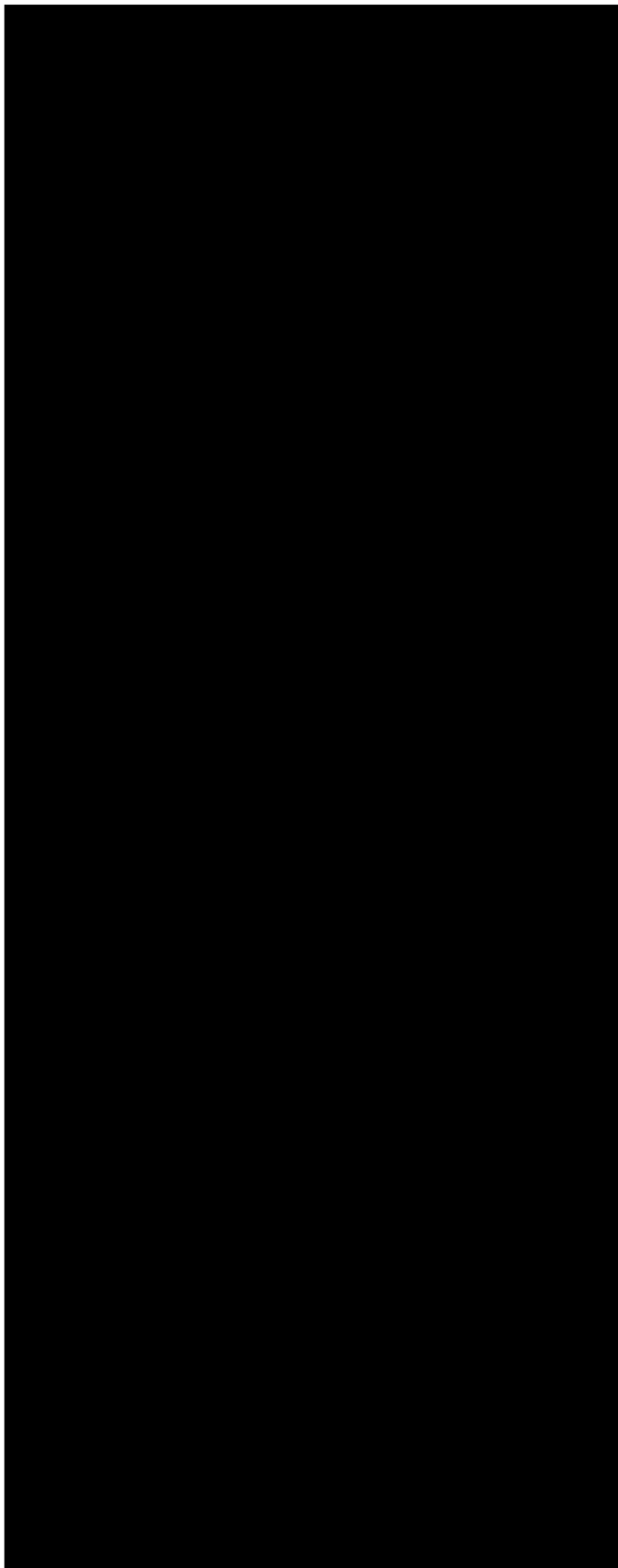
1   FAC ¶¶ 106-07.

2   The other two examples made clear that Plaintiffs ventured far outside any ordinary use of

3   Copilot to manufacture them, and even then failed to get Copilot to produce an actual copy. FAC

4   ¶¶ 100-04, 113-26. Plaintiffs offered no example of a verbatim copy of code.

5   This Court granted in part and denied in part GitHub and Microsoft's motions to dismiss

6   the First Amended Complaint. ECF No. 189. On Article III standing, the Court noted that its

7   prior order found that Plaintiffs had standing to seek injunctive relief because "they plausibly

8   alleged that there is 'at least a substantial risk that Defendants' programs will reproduce

9   Plaintiffs' licensed code as output' in the future." ECF No. 189 at 6 (quoting ECF No. 95 at 9).

10 The Court again held that all Plaintiffs had standing to seek injunctive relief. ECF No. 189 at 7-9.

11 Regarding standing to seek monetary damages, the Court held that because the "first amended

12 complaint … now include[d] examples of licensed code owned by Does 1, 2, and 5 that has been

13 output by Copilot," it concluded that "Does 1, 2, and 5 have adequately alleged 'particular

14 personalized injury' sufficient to confer standing for monetary damages." ECF No. 189 at 6. But

15 because "Does 3 and 4 have again failed to raise instances in which their code was output by

16 Copilot," the Court dismissed with prejudice Does 3 and 4's request for damages. ECF No. 189

17 at 6, 9.

18   As to Defendants' motions to dismiss under Rule 12(b)(6) for failure to state a claim, the

19 Court concluded that Plaintiffs' state law claims were preempted by Section 301 of the Copyright

20 Act. ECF No. 189 at 9-14. It dismissed Plaintiffs' intentional and negligent interference with

21 prospective economic relations, unjust enrichment, unfair competition, and negligence claims

22 with prejudice. ECF No. 189 at 9, 16.

23   Turning to Plaintiffs' claims under § 1202(b)(1) and (b)(3) of the DMCA, this Court

24 observed that "'[c]ourts have held that no DMCA violation exists where the works are not

25 identical.'" ECF No. 189 at 15 (quoting *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search*

26 *Optics, LLC*, No. 22-cv-1186-TWR-BLM, 2023 WL 3366534, at *12 (S.D. Cal. May 9, 2023)).

27 The Court concluded that Plaintiffs did not plead identical works, because "the examples

28 Plaintiffs provide[d] with respect to Does 1, 2, and 5 state that the Copilot output is a 'modified

1   format,' 'variation[],' or the 'functional[] equivalent' of the licensed code." ECF No. 189 at 15

2   (quoting ECF No. 97-3 ¶¶ 103, 110, 120). The Court concluded that such pleading failed

3   § 1202(b)'s identicality requirement. ECF No. 189 at 16. Noting that it was "unlikely that this

4   deficiency could be cured by the allegation of additional facts," this Court dismissed Plaintiffs'

5   § 1202(b)(1) and (b)(3) claims, but "out of abundance of caution" granted leave to amend. *Id.*

6   **C.   Plaintiffs' New Allegations.**

7   In the Second Amended Complaint, three claims remain. Counts two and three re-allege

8   Plaintiffs' breach of contract claims without modification. SAC ¶¶ 236-51 (Count 2); SAC

9   ¶¶ 252-62 (Count 3). Additionally, despite the Court's dismissal of all state law tort claims with

10   prejudice, Plaintiffs continue to plead for monetary relief in the form of unjust enrichment and

11   punitive damages. Prayer ¶ 263(e).

12   The changes in the Second Amended Complaint all concern Plaintiffs' attempt to plead

13   some viable claim for violation of § 1202(b)(1) and (b)(3) of the DMCA. SAC ¶¶ 204-35 (Count

14   1). To begin with, Plaintiffs remove from their Complaint any reference to or request for

15   damages under those provisions. *See* ECF No. 197-4 at 57, 68-69. This appears to be a

16   concession that, as this Court anticipated, Plaintiffs cannot plausibly allege that Copilot has ever

17   actually emitted an identical copy of one of their copyrighted works without the associated CMI.

18   Instead, Plaintiffs attempt to allege that "there is a substantial risk" that this could happen. SAC

19   ¶ 206.

20   In support of these claims, Plaintiffs continue to allege generally that they "published

21   Licensed Materials they owned a copyright interest in to at least one GitHub repository under one

22   of the Suggested licenses." SAC ¶¶ 19-23. They assert that because Copilot was trained on

23   public GitHub repositories, they can be "reasonably certain" their Licensed Materials were

24   "ingested by Copilot and [are] sometimes returned to users as Output." SAC ¶ 109. The Second

25   Amended Complaint retains the same examples using Does 1, 2, and 5's code, as well as the same

26   allegations that Copilot's output is a "modified format," "variation[]," or the "functional[]

27   equivalent" of their code. SAC ¶¶ 115, 122, 124, 130, 132. And it alleges that GitHub provides a

28   "user-settable Copilot filter called 'Suggestions matching public code,'" which allows users to

1   block matching "excerpts of 'about 150 characters.'" SAC ¶ 145.

2   Plaintiffs also continue to assert that "about 1% of the time, a suggestion may contain

3   code snippets longer than ~150 characters that matches code from the training data." SAC

4   ¶¶ 102, 207. From this, Plaintiffs suggest that "Defendants have … breached Plaintiffs' and the

5   Class's Licenses 12,000 times," based on the fact that Copilot had 1,200,000 users in June 2022,

6   SAC ¶ 103—asserting, with zero support, that *every* work contained in a public repository has a

7   1% chance of having some snippet of it being output. Plaintiffs elsewhere contradict this

8   assertion, however, by alleging that "Copilot returns the solution it has found in the *most*

9   *projects*." SAC ¶ 91 (emphasis added).

10   Plaintiffs' attempt to cure the deficiency in the First Amended Complaint rests on two

11   new sets of allegations. The first is an expanded description of GitHub's duplication-detection

12   feature. SAC ¶¶ 146-54. The duplication-detection tool allows users to "check[] code -

13   completion suggestions with their surrounding code of about 150 characters against public code

14   on GitHub." SAC ¶ 146. If a user has decided to "block suggestions matching public code,"

15   then, "[i]f there is a match, or a near match, the suggestion is not shown[.]" SAC ¶ 146. If, on

16   the other hand, the user has decided to "allow[] suggestions that match public code," then Copilot

17   provides "details about the matching code." SAC ¶ 147. Plaintiffs rely on the existence of the

18   duplication-detection tool as demonstrating that Copilot suggestions sometimes contain short

19   passages of identical code.

20   The second set of new allegations concerns "[r]ecent academic research shows that the

21   likelihood Plaintiffs' or class members' code would be emitted verbatim is only increasing."

22   SAC ¶ 104. Plaintiffs' sole citation is to an article first submitted two years ago. SAC ¶ 104 n.20

23   (citing Nicholas Carlini, et al., Quantifying Memorization Across Neural Language Models,

24   arXiv (submitted Feb. 15, 2022, revised Mar. 6, 2023), https://arxiv.org/pdf/2202.07646.pdf).

25   Plaintiffs allege that this article shows "that Copilot is more likely to emit duplicates of

26   memorized training data as it continues to scale." SAC ¶ 106.

27

28

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

ARGUMENT

I.   PLAINTIFFS FAIL TO STATE A DMCA CLAIM.

Plaintiffs' latest attempt to assert a claim under § 1202(b)(1) and (b)(3) of the DMCA still fails to state a claim. Section 1202(b)(1) prohibits a defendant from "intentionally remov[ing] or alter[ing] [CMI]," *see* SAC ¶ 205, while § 1202(b)(3) bars "distribut[ing] … copies of *works …* knowing that [CMI] has been removed or altered," *see* SAC ¶ 206 (emphasis added). CMI is defined as information "conveyed in connection *with copies* … of a *work*"—that is, in or directly connected with "the body of or the actual *work itself*"—and not a portion of the work. *Mills v. Netflix, Inc.*, No. 19-cv-7618-CMB, 2020 WL 548558, at *2-3 (C.D. Cal. Feb. 3, 2020) (emphases added) (first part quoting 17 U.S.C. § 1202(c)). But all Plaintiffs continue to plead in the Second Amended Complaint is alleged removal of CMI with a proffered *snippet* that may incidentally match *a portion* of works—works that Plaintiffs fail to identify.

The claim fails for multiple reasons. As with the prior amendment, in attempting to solve one defect in their DMCA claim, Plaintiffs lay others bare. Here, Plaintiffs' remaining matching-snippet theory fails because they never allege either actual or threatened removal of CMI from an identified work. Part I.A., *infra*. In any event, as this Court anticipated, Plaintiffs have failed to identify removal of CMI from Copilot output that is identical to code over which they claim copyright ownership. Part I.B., *infra*. And Plaintiffs' Complaint now makes abundantly clear, even on the pleadings, that Plaintiffs will never be able to allege that any removal of CMI makes copyright infringement easier to commit or harder to detect. Part I.C., *infra*.

    A.   <u>Plaintiffs Fail To Identify Any Copyrighted Work From Which CMI Has Been Or Is Likely To Be Removed.</u>

Courts have consistently required plaintiffs to identify copyrighted works from which CMI has been or will be removed. *See Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1175 (N.D. Cal. 2019) (dismissing DMCA claim where plaintiff failed to plead "any facts to *identify which* photographs had CMI removed") (emphasis added); *see also Kadrey v. Meta Platforms, Inc.*, No. 23-CV-03417-VC, 2023 WL 8039640, at *2 (N.D. Cal. Nov. 20, 2023) (dismissing DMCA claim where there were "no facts to support the allegation that [the Large

1   Language Model Meta AI] ever distributed the *plaintiffs'* books") (citing *Menzel*, 390 F. Supp. 3d

2   at 1175) (emphasis added). Moreover, a bare assertion of a copyright interest "does not amount

3   to ownership of a copyright in the material," and a pleading that does "not sufficiently allege[]

4   ownership of a copyright in a work" must be dismissed. *Balik v. Spiral Toys Inc.*, No. 15-cv-

5   8112-GW(PLAX), 2016 WL 7496134, at *3 (C.D. Cal. Feb. 1, 2016); *see id.* (citing *Feist Pubs.,*

6   *Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 361 (1991)). Plaintiffs here continue to vaguely assert

7   that their copyrighted works were in Copilot's training set, pointing to the same code sections

8   posted by Does 1, 2, and 5, while not identifying any code from Does 3 and 4. *See* SAC ¶¶ 113,

9   117, 125. And Plaintiffs continue to plead in a conclusory manner that "they owned a copyright

10   interest in" the code that they posted. SAC ¶¶ 19-23.

11   Plaintiffs' examples in the Second Amended Complaint, repeated verbatim from the First

12   Amended Complaint, show how broad—and legally untenable—Plaintiffs' theory is. In the

13   example pertaining to Doe 2, Plaintiffs caused Copilot to emit a list of the ▆▆▆▆▆▆▆▆▆

14   ▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆. SAC ¶¶ 112-

15   116. Plaintiffs do not allege that this list of ▆▆▆▆▆ constitutes a full, identical copy of a

16   work in which they have an interest. Nor could they claim rights in the allegedly reproduced

17   aspects of it, since this "code" is merely a table of constants reflecting ▆▆▆. *See Kevin Barry*

18   *Fine Art Assocs. V. Ken Gangbar Studio, Inc.*, 391 F. Supp. 3d 959, 966 (N.D. Cal. 2019) (citing

19   *Feist*, 499 U.S. at 344-45). Although "a sufficiently original compilation of those facts" could be

20   copyrightable, *id.*, here Plaintiffs do not even claim *that*, admitting that "[t]he ▆▆▆▆▆ are *not*

21   [arranged] in any semantically significant order." SAC ¶ 113 (emphasis added). Yet Plaintiffs

22   seem to be suggesting that if Doe 2 is not credited as the author of this obviously uncopyrightable

23   code snippet, Defendants have violated § 1202(b).

24   The same problems extend to Doe 1's code "▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆ pieces,"

25   and, likewise, there are no allegations explaining how Doe 5's code "▆▆▆▆▆▆▆▆▆

26   ▆▆▆▆▆▆▆▆▆" could be copyrightable. SAC ¶¶ 118, 126. Plaintiffs thus nowhere

27   identify any copyrighted work owned by them or any registration of such work. This underlying

28   defect has only been made clearer in repeated rounds of amendments, and it dooms Plaintiffs'

1   DMCA claim. Especially where Plaintiffs' entire theory is premised on some purported

2   *likelihood* of future output, it is incumbent upon Plaintiffs to provide the copyrighted work from

3   which they claim CMI will be removed.

4   Moreover, at this stage of the litigation, it is abundantly clear that Plaintiffs' case rests on

5   alleged outputs of short *snippets* of code, and not outputs of a complete work. Courts have

6   rejected CMI removal liability based on mere excerpts of works, rather than complete works. For

7   instance, courts have rejected mere "framing" of a photograph to not include CMI, *Falkner v.*

8   *Gen. Motors LLC*, 393 F. Supp. 3d 927, 938-39 (C.D. Cal. 2018); excerpting lecture notes and

9   study questions from textbooks without reproducing CMI, *Faulkner Press, L.L.C. v. Class Notes,*

10   *L.L.C.*, 756 F. Supp. 2d 1352, 1356, 1359 (N.D. Fl. 2010); copying "aspects" of architectural

11   works but "omitting" the plaintiff's CMI, *Design Basics, LLC v. WK Olson Architects, Inc.*, No.

12   17-cv-7432, 2019 WL 527535, at *5 (N.D. Ill. Feb. 11, 2019); *Frost-Tsuji Architects v. Highway*

13   *Inn, Inc.*, No. 13-cv-00496-SOM, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015), *aff'd*, 700 F.

14   App'x 674 (9th Cir. 2017); or incorporating the underlying content from the original copy into

15   some different form or distinct work without CMI, *Kelly v. Arriba Soft. Corp.*, 77 F. Supp. 2d

16   1116, 1122 (C.D. Cal. 1999) (thumbnail versions of images). Those cases are on all fours here.

17   And indeed, they are all the more salient because Plaintiffs' attempt to overcome their identicality

18   defect relies on Copilot's duplication-detection tool, which by definition only applies to snippets.

19   Snippet-based allegations get Plaintiffs nowhere.

20   At bottom, Plaintiffs' claims are based on the abstract allegation that some unidentified

21   snippet from some unidentified work has some undefined likelihood of being transposed to

22   another context, purportedly without CMI that accompanies the work as a whole. It makes no

23   sense that such a claim would be cognizable in light of § 1202's aims. That statute was enacted

24   to create so-called paracopyright protections, "assist[ing] in tracking and monitoring uses of

25   copyrighted works, as well as licensing of rights and indicating attribution, creation and

26   ownership." S. Rep. No. 105-190 (1998) at 16. That is why CMI is defined not as any

27   information conveyed with any content, but specifically as "information conveyed in connection

28   with copies … of a work." 17 U.S.C. § 1202(c). Were it otherwise, what was intended as a

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
No. 4:22-CV-6823-JST

1   narrow paracopyright protection would swallow copyright protection whole, allowing plaintiffs to

2   plead "removal" of CMI from mere snippets of an unidentified work that may have no claim to

3   any copyright protection at all.

4   This fundamental defect is endemic to Plaintiffs' theory of § 1202(b) liability. As the

5   Second Amended Complaint acknowledges, Copilot's output is limited to short "passages" of

6   code representing suggested completions for coding processes initiated by a user. SAC ¶ 102.

7   Copilot's output will therefore virtually never constitute something that could be described as a

8   copy of a work, let alone an *identical* copy. As the case law above demonstrates, § 1202(b) is

9   about identical "copies … of a work"—not about stray snippets and adaptations. This Court

10 should again dismiss Plaintiffs' § 1202(b)(1) and (b)(3) claims, because Plaintiffs have not

11 identified any copyrighted works and have only pleaded snippets of works.

12   **B.**   **Plaintiffs' DMCA Claim Fails § 1202(b)'s Identicality Requirement.**

13   Plaintiffs' 1202(b)(1) and (b)(3) claims fail for a related but independent reason. This

14 Court previously held that the Complaint failed to allege the removal of CMI from an identical

15 copy of the underlying work. ECF No. 189 at 16. Nothing about the revised pleadings changes

16 this. It is hard to see how a complaint that failed to allege a likelihood of identical copies could

17 improve by noting a tool that *blocks* them—the duplication-detection feature obviously makes it

18 even *less* plausible that an identical snippet of Plaintiffs' code would be emitted.

19   As this Court held in dismissing the First Amended Complaint, "'[c]ourts have held that

20 no DMCA violation exists where the works are not identical.'" ECF No. 189 at 15 (quoting

21 *Advanta-STAR*, 2023 WL 3366534, at *12). It went on to observe that, "'[e]ven where the

22 underlying works are similar, courts have found that no DMCA violation exists' unless the works

23 are *identical*." ECF No. 189 at 15 (quoting *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. cv-

24 20-1931-DMG, 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020) (emphasis added); and citing

25 *Frost-Tsuji*, 2015 WL 263556, at *3 (finding no § 1202(b) violation where the allegedly

26 infringing drawing was "not identical")).

27   To begin with, Plaintiffs now *concede* they cannot allege that Copilot has ever actually

28 generated an identical match with work in their repositories. Moreover, as with the First

1   Amended Complaint, the Second Amended Complaint continues to acknowledge that "more

2   often" Copilot's suggestions are "a modification," SAC ¶ 108—a concession on the lack of

3   identicality that, as this Court concluded, meant that Plaintiffs "ha[d] effectively pleaded

4   themselves out of their Section 1202(b)(1) and 1202(b)(3) claims." ECF No. 189 at 14. So

5   instead, Plaintiffs retreat to the position that maybe, someday, Copilot could emit a snippet of

6   matching text that is identical to some passage of Plaintiffs' text—despite the fact that even

7   Plaintiffs' own deliberate, acrobatic prompting could not cause it to do so.

8   At best, however, Plaintiffs' newest allegations support only the barest hypothetical

9   possibility of identical copies. Plaintiffs first point to Copilot's duplication-detection feature.

10   The duplication-detection tool allows Copilot users to decide whether to block a match of around

11   150 characters of code. SAC ¶ 146. Plaintiffs' theory is that a user might elect not to use the

12   blocking feature, but instead "review[] the references" it identifies and "decide how to proceed."

13   SAC ¶ 150 (emphases omitted). Only at the end of this series of conditional events could

14   Plaintiffs plausibly posit that a user could view an identical match of around 150 characters of

15   their code and proceed to use it without attribution. This chain of conjectural possibilities fails to

16   "'nudge[] [Plaintiffs'] claim[]' … 'across the line from conceivable to plausible.'" *Ashcroft v.*

17   *Iqbal*, 556 U.S. 662, 680 (2009) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)).

18   And in any event, the mere existence of such a feature does not make it more likely that Copilot

19   would ever output an identical copy of *Plaintiffs*' works.

20   Plaintiffs' reliance on a two-year-old academic paper fares no better. Plaintiffs invoke it

21   to support their claim that "the likelihood Plaintiffs' or class members' code would be emitted

22   verbatim is only increasing." SAC ¶ 104 (citing Nicholas Carlini et al., *Quantifying*

23   *Memorization Across Neural Language Models* at 6 (Mar. 6, 2023),

24   https://arxiv.org/pdf/2202.07646.pdf (hereinafter "Carlini")). But this assertion goes no further

25   than the deficient allegations from Plaintiffs' last complaint. It certainly does not rehabilitate

26   Plaintiffs' own concession that, still, "more often," Copilot's suggestions are "a modification."

27   SAC ¶ 108.

28   The academic paper also fails to move the needle because it does not say what Plaintiffs

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

1   insinuate it does. Plaintiffs' quoted passage actually states that verbatim matches occur only

2   when "prompted appropriately," SAC ¶ 104 (quoting Carlini at 1). It explains that "the GitHub

3   Copilot model … reportedly rarely emits memorized code in benign situations, and most

4   memorization occurs only when the model has been prompted with long code excerpts that are

5   very similar to the training data." Carlini, *supra*, at 6. Plaintiffs tried to do exactly that in their

6   last complaint, deploying every trick they could think of to "prompt" Copilot "appropriately" so

7   as to generate an identical copy of *their* code and they *still* were not able to. It does them no good

8   to cite a paper that suggests the bare possibility that someone else could theoretically—making

9   something other than "benign" use of Copilot—generate a match to someone else's code. So all

10   the newly added Carlini paper does is confirm the very defect that led to dismissal of the First

11   Amended Complaint.

12   Plaintiffs speculate that "as generative AI models such as Copilot increase capacity and

13   continue to scale, it becomes more likely that training data will become memorized and emitted

14   verbatim." SAC ¶ 105 (emphasis omitted). At most, this might suggest that some future version

15   of Copilot with larger capacity may be more likely to output Plaintiffs' code verbatim. SAC

16   ¶ 106. But Plaintiffs fail to account for the fact that, even if Copilot launched with a model with

17   larger capacity, the GitHub public repository could also expand and "scale" in proportion. That

18   would result in more data to train on by a corresponding percentage, negating Plaintiffs'

19   suggestion of increased matches with existing training material. Thus, there is nothing behind

20   Plaintiffs' bare assertion that verbatim matches will become more likely.

21   That leaves only Plaintiffs' steady refrain, already deemed insufficient, that "GitHub has

22   admitted that about 1% of the time, a suggestion may contain code snippets longer than ~150

23   characters that matches code from the training data." SAC ¶ 207; *see also* SAC ¶¶ 102-03. Let's

24   be very clear about what this statement *does not* mean: It is *not* saying that there is a 1%

25   likelihood for every work contained in the training data that some 150-character snippet of that

26   work will be a Copilot suggestion, as Plaintiffs appear to insinuate. SAC ¶ 103. It means only

27   that 1% of Copilot's outputs will contain a snippet matching *something* in the program's vast

28   training set. If you surveyed a random sample of people on their favorite books, perhaps 1% of

1   responses would name a science fiction novel; that does not mean that every science fiction novel

2   ever published has a 1% chance of being named. Why not? Because some sci-fi novels are very

3   popular and some are not. As the Complaint concedes, Copilot's outputs work the same way,

4   affording high predictive weight to code that is "found in the most projects." SAC ¶ 91. There

5   are only so many ways, for example, to convert Fahrenheit to Celsius, or to calculate the speed of

6   a falling object, or to find the prime factors of a large number. No allegations in the Complaint

7   support any inference that *Plaintiffs'* code has anything close to even a 1% chance of being output

8   as an identical snippet of ~150 characters.

9   **C.   Plaintiffs Cannot Allege The Required Likelihood Of Infringement.**

10   The Court should also dismiss on the independent ground that the Second Amended

11   Complaint does not allege any likelihood that Copilot will result in copyright infringement of any

12   of Plaintiffs' works, because the new allegations of duplicate-blocking, coupled with new

13   precedent from this Court, further compel dismissal on this basis. *See Tremblay v. OpenAI, Inc.*,

14   Nos. 23-cv-03223, 23-cv-03416, 2024 WL 557720, at *4 (N.D. Cal. Feb. 12, 2024) (dismissing

15   § 1202 claim where complaint failed to show that the alleged tampering with CMI "knowingly

16   enable[d] infringement").

17   To plead a claim under § 1202(a) pertaining to false CMI, a plaintiff must allege that the

18   defendant acted "with intent to induce, enable, facilitate, or conceal infringement." A claim

19   under § 1202(b) regarding altered or removed CMI similarly requires that a defendant either have

20   "know[ledge]" or have "reasonable grounds to know" of the same result. Of necessity, then, the

21   tampering with CMI contemplated by § 1202 must be part of a scheme involving purposeful

22   infringement. *Stevens v. Corelogic, Inc.*, 899 F.3d 666 (9th Cir. 2018), recognizes this

23   requirement and provides that a § 1202 claim requires a showing "that future infringement is

24   *likely* … to occur as a result of the removal or alteration of CMI." *Id*. at 675 (emphasis added).

25   Beyond threadbare allegations (*e.g.*, SAC ¶¶ 221, 223), Plaintiffs still cannot explain how

26   infringement is likely to occur in the context of their DMCA claims. Although Plaintiffs assert

27   that GitHub and Microsoft's alleged tampering with CMI "facilitate[s] or conceal[s]"

28   infringement because Plaintiffs "are prevented from knowing or learning" if their code has been

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST

1   outputted, SAC ¶ 222, there is no realistic probability that such outputs will actually occur—and

2   even less of a likelihood that they would amount to infringement. Judge Martínez-Olguín

3   recently dismissed a § 1202 challenge to a generative AI model on this basis. *See Tremblay*, 2024

4   WL 557720, at *4 (rejecting the notion that removing CMI enables purposeful infringement

5   "because ChatGPT users will not know if any output is infringing" and dismissing § 1202 claim

6   under *Stevens*). Worse, the SAC expressly describes a feature in Copilot that would dispel

7   potential matches, the duplication-detection feature, making it clear from the face of the SAC that

8   it is highly implausible Copilot was designed to facilitate infringement.

9   As explained above (at 16-17), the chances that the small portion of Copilot's outputs

10   containing short snippets that match something in Copilot's enormous training set would include

11   Plaintiffs' code at all are vanishingly small. The chances that the outputting of those snippets

12   infringe Plaintiffs' copyright—much less purposefully so—are virtually zero. Even if a snippet

13   matched Plaintiffs' code, Plaintiffs would have to overcome numerous copyright-law obstacles to

14   establish infringement, including the copyrightability of such stray snippets in light of doctrines

15   of merger and scènes à faire, the requirement of virtually identical copying of thinly protected

16   elements, the scope of authorization through licensing, and fair use. Because the Second

17   Amended Complaint contains no plausible allegations that Copilot's operation with respect to

18   CMI is likely to result in infringement, the § 1202(b) claims should be dismissed.

19   **II.   PLAINTIFFS' REQUESTS FOR UNJUST ENRICHMENT AND PUNITIVE**
     DAMAGES SHOULD BE DISMISSED.

20

21   The Court previously dismissed with prejudice Plaintiffs' unjust enrichment claim as

22   preempted by the Copyright Act. ECF No. 189 at 11-13. Plaintiffs nevertheless continue to plead

23   a request for monetary relief in the form of unjust enrichment. SAC ¶ 263(e). Since the DMCA

24   claim is expressly limited to injunctive relief, this request can only be predicated upon breach of

25   contract claims. As for those claims, although unjust enrichment is sometimes a "theory

26   underlying a claim that a defendant has been unjustly conferred a benefit," a plaintiff must plead

27   "'mistake, fraud, coercion, or request.'" *Astiana v. Hain Celestial Grp., Inc.*, 783 F.3d 753, 762

28   (9th Cir. 2015) (quoting 55 Cal. Jur. 3d Restitution § 2). None of Plaintiffs' remaining state law

1   claims contains any such allegation. *See* SAC ¶¶ 236-62. Thus, even assuming unjust

2   enrichment were an available remedy for breach of contract claims that otherwise seek ordinary

3   damages, the standard for such relief is not met by the Second Amended Complaint. To the

4   extent Plaintiffs' continuing request for unjust enrichment monetary relief stems from the claim

5   of unjust enrichment that was previously dismissed with prejudice, then GitHub and Microsoft

6   move to strike that request as improper. *See Santa Clara Valley Water Dist. v. Olin Corp.*, No.

7   07-cv-03756-RMW, 2007 WL 2890390, at *5 (N.D. Cal. Sept. 28, 2007) ("Improper prayers for

8   relief are proper subjects for a motion to strike."); *see also Howard v. Blue Ridge Bank*, 371 F.

9   Supp. 2d 1139, 1146 (N.D. Cal. 2005) (similar).

10   The punitive damages request should be dismissed as well. Plaintiffs continue to plead

11   punitive damages, even though this Court dismissed with prejudice any state law tort claims that

12   could have supported such a request. *See* SAC ¶ 263(e). The only state law claims left are for

13   breach of contract. Under California law, "punitive damages are never recoverable in routine

14   breach of contract cases." *Oliver v. Astrazeneca Pharms., LP*, No. 10-cv-03073-RGK-AJWX,

15   2011 WL 13214269, at *10 (C.D. Cal. Mar. 25, 2011) (quoting *Power Standards Lab, Inc. v. Fed.*

16   *Express Corp.*, 127 Cal. App. 4th 1039, 1047 (2005)); *see Foley v. Interactive Data Corp.*, 47

17   Cal. 3d 654, 698-700 (1988) (no punitive damages available for breach of contract). The prayer

18   for monetary relief in the form of unjust enrichment and request for punitive damages should be

19   dismissed.

20   **CONCLUSION**

21   Plaintiffs have had three opportunities to plead their claims. As the foregoing discussion

22   demonstrates, the § 1202 claim and the requests for unjust enrichment and punitive damages are

23   unsupportable. The Court should now dismiss with prejudice.

24

25

26

27

28

1

2 Dated: February 28, 2024   Orrick, Herrington & Sutcliffe LLP

3

4   By: _____/s/ *Annette L. Hurst*_____

5   ANNETTE L. HURST
                                                                Attorneys for Defendants
6   GitHub, Inc. and Microsoft Corp.

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

GITHUB AND MICROSOFT'S MTD SECOND
AMENDED COMPL.
NO. 4:22-CV-6823-JST